International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 3, No. 1, 54-69

# A Genetic Approach Wrapped Support Vector Machine for Feature Selection Applied to Parkinson's Disease Diagnosis

Ayoub Bouslah and Nora Taleb

LRI Laboratory, Department of Informatics, Badji-Mokhtar Annaba University, P.O. Box 12, 23000, Annaba, Algeria
ic.bouslah@gmail.com and talebnr@hotmail.fr

**Abstract.** Parkinsons disease (PD) is found to be a challenging issue which can offer a computerized estimate about classification of PD to patient people and healthy for normal people. Due to the importance of that problem, several types of biomedical data can be analyzed to accurately detect PD by using different learning methods. This work considers the diagnosis of PD based on voice data by using non-linear support vector machine (SVM). However SVM is known as the one of the fast and accurate learning methods, selection of relevant feature elements of PD dataset can be effective on improving the classification performance of SVM. To this end, this paper proposed an SVM in parallel with GA based feature reduction model for selecting the most relevant features to get Parkinson's disease. The GA-SVM resulted in improved accuracy, sensitivity and area under curve (95%, 98% and 92% respectively) compared to the other learning methods and feature selection algorithms. The GA-SVM provides a better, more accurate identification for presence of vocal disorder from speech recordings leading to more timely diagnosis.

**Keywords:** Parkinsons disease (PD) · Feature selection · Genetic algorithm (GA)· Classification · Support Vector Machine (SVM).

## 1 Introduction

Parkinson's disease is the second most neuron destructive disease after Alzheimer's disease. Although its underlying cause is unknown, the symptoms associated with PD may be significantly lessened if they are detected in the early stages of the disease [1, 2]. PD is characterized by tremors, rigidity, slow motion, asymmetry of motor symptoms and impaired posture [3, 4]. A person with Parkinson's will progressively lose their physical abilities and become worse if there is no health care or appropriate solution. This disease occurs in all races and spreads in one to two individuals in a thousand. Its spread increases with age. It has been estimated that around 40% of people with this disease may not be diagnosed [5]. In recent years, the efforts to understand and characterize PD have intensified, a number of data mining and machine learning algorithms are developed to predict the early stage of Parkinson's disease from biomedical data using voice, gait, and wearable sensors [6–8]. Voice signal recording is the earliest, easiest and non-invasive technique for diagnosis of PD [9]. As most people with PD suffer from speech disorders [10], it could be considered a very reasonable way for detecting PD [11, 12]. In many cases, data mining and machine learning algorithms are used on a very large scale of Parkinsons disease data [13, 14]. This would create high computational complexity and lower efficiency. In order to overcome this problem, a range of feature selection algorithms are developed to identify the most significant features for predicting Parkinson's disease [15, 16]. The most recent attempts at diagnostic improvement consider the feature selection of the voice data set as in Kaya et al. (2011) [17] and A. Tsanas et al. (2012) [18], the classification methods as in Sakar (2013) [19] and M. Can (2013) [20] or both of them as in (K. Shahsavari et al., 2016; S. Yanget al, 2018) [21, 22]. The results show that feature selection process is very important to improve diagnostic accuracy. Subsequently, this paper presents an attempt to improve the diagnosis of Parkinsons disease. The main contributions of this work are proposing a new genetic algorithm based feature selection coupled with SVM in the Parkinsons diagnosis problem. In addition, we present a comparison of various types of feature selection algorithms. More generally, feature selection algorithms are classified into the following types, information gain, T-test, CFS, RFS (R-value), relief, MRMR because they are fast and efficient [40]. In addition, genetic algorithms are inductive. Adaptive random search techniques make it possible to exploit information accumulated on an unknown search space and then search for promising new subspaces [23]. Finally, the SVM-based RFE-CBR is a wrapper feature selection algorithm that uses criteria derived from the coefficients in original SVM models to assess features. It recursively removes features that are not informative. Compared to other wrapper techniques, SVM-based RFECBR does not use the precision of cross-validation on the training set as a selection criterion. As a result, it is less subject to overfitting and remains fast even though the original feature set is large [24]. The remainder of the paper is organized in the following manner. Section 2 contains related work. Section 3 introduces the methodology. Section 4 presents the data and the efficiency of our method are compared to other methods and finally the conclusion of the work.

## 2 Related work

Using the speech samples for the diagnosis of PD has been the subject of several investigations. For instance, Shahbaba et al. (2009) [24] used a non-linear model based on Dirichlet mixtures for the diagnosis of PD. An 87.7% classification accuracy was obtained with this method. Little et al. (2009) [25] conducted a remarkable study about PD identification, they employed a Support Vector Machine (SVM) classifier with Gaussian radial basis kernel functions to predict PD, and also performed feature selection to select the optimal subset of features from the whole feature space, and the best accuracy rate of 91.4% was obtained by the best model. Das (2010) [26] carried out a comparative study of artificial neural networks (ANN), DMneural, regression and decision trees for the diagnosis of the PD using speech samples.

The experimental results showed that the ANN method achieved a 92.9% general classification performance. Guo et al. (2010) [27] proposed a hybrid model based on expectation maximization (EM) and GA, the classification accuracy obtained is 93.1% . Ozcift and Gulten (2011) [28] combined the CFS (correlation based feature selection) algorithm with the rotation forest classifiers (RF) of 30 machine learning algorithms to identify the PD, and the best classification accuracy 87.13 % was achieved by the proposed CFS-RF system. Chen et al. (2013) [29] used the feature reduction method to exclude redundant information from the original PD speech signal, embedded in the fuzzy classifier for PD diagnosis. They achieved an average classification accuracy of 96.07%. In another study (2016) [30], the authors have also proposed using Extreme Learning Machine (ELM) and Extreme Kernel Learning Machine (KELM) for early diagnosis of PD. Experimental results showed that the proposed KELM method combined with the feature selection method provides very promising classification accuracy with a maximum accuracy of 96.47% and an average accuracy of 95.97% over 10-fold CV.

More recently, Peker et al. (2015) [31] proposed to combine a maximum redundancy maximum relevance attribute selection algorithm with the complex-valued artificial neural network to detect PD; the classification accuracy of 98.12% was obtained by the proposed methodology. Lahmiri et al (2019) [32] focused on evaluating the performance of eight pattern ranking techniques, including Battacharyya, GA, ROC, RFECBR, Wilcoxon, Entropy, t-test, and MI coupled to a nonlinear support vector (SVM) machine to distinguish patients with Parkinson's disease from healthy control subjects. The core parameters of the SVM classifier's kernel rbf were optimized using the Bayesian optimization technique. The results obtained show that the classifier obtained the highest classification accuracy (92.21%) with 14 vocal features identified by the pattern ranking technique based on the Wilcoxon method. The highest specificity (82.79%) when formed with the first 13 significant voice models identified by the ROC based attribute selection technique. The highest sensitivity, 99.63%, with a single voice pattern under the ROC based attribute selection technique.

## 3 Proposed framework for feature selection

This study proposed a SVM in parallel with GA based feature reduction model for selecting the most relevant features to get Parkinson's disease. The overall architecture is illustrated in **Fig. 1**, in which each stage is described down.

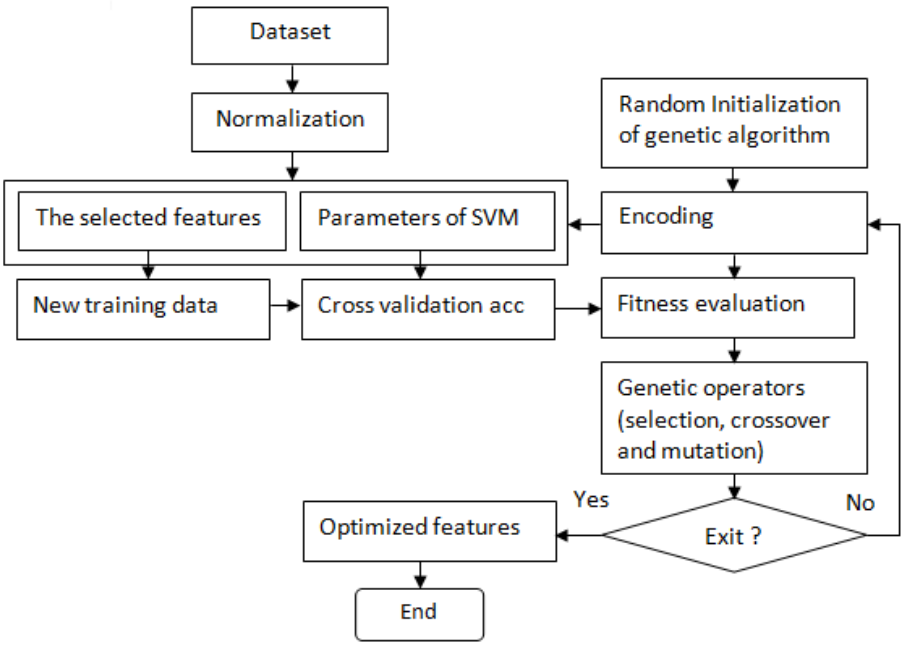


**Fig. 1.** The overall architecture

### 3.1 Genetic algorithm based feature selection

GA being a heuristic search algorithm [33], it is commonly used to identify relevant features for high dimensional datasets. A wrapper-based GA is used in this paper to select the most relevant features. The essential role of the wrappers used in the GA is to find the space and evaluate each subset by running a model. The search used in the GA creates a set of candidate solutions (individual) and the optimization problem is established for an improved solution. **Fig. 2** shows the working principle of a GA [34]. In genetic algorithms, a potential solution to problem is encoded as a chromosome. This group of chromosomes, i.e. population, is the search space of the algorithm. A fitness function is used to evaluate performance of each chromosome to measure its closeness to solution. The initial chromosomes as parents are used to generate new offspring chromosomes with

the use of genetic operators namely selection, mutation and crossover. At each turn, the number of chromosomes in the population is retained constant with eliminating chromosomes having lowest fitness value. This process is repeated until the solution is found or the maximum number of iterations is reached [35]. We designed a GASVM algorithm for identifying significant features of Parkin-

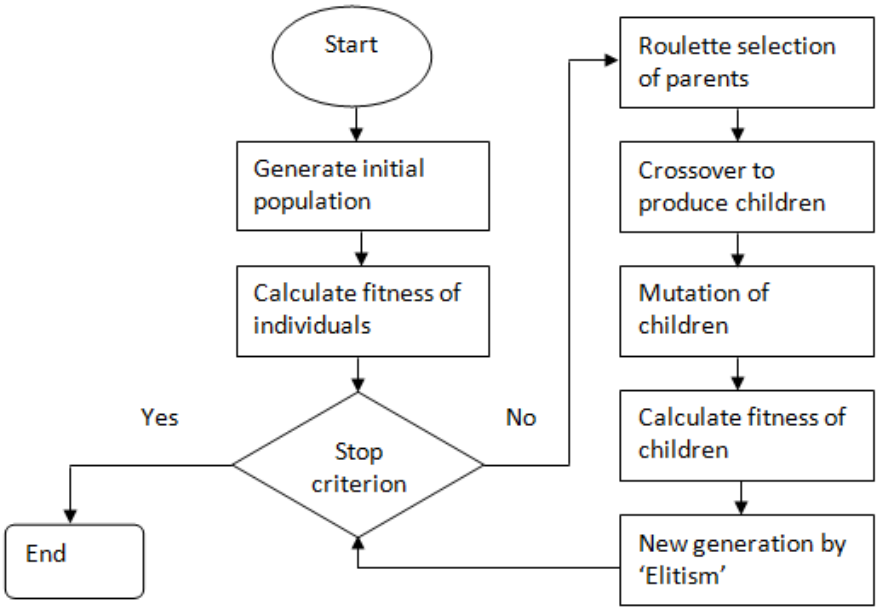


**Fig. 2.** The working principle of genetic algorithm [34]

son's diseases diagnosis and the details of the proposed algorithm is given as follows.

**Encoding** In our model, chromosomes (features representing the specification of a patient) are encoded as bit strings of 1 and 0. In this scheme, 1 represents the selection of a feature and 0 a nonselection. The number of genes in the chromosome is 22, which corresponds to the size of the features of PD dataset.

**Random Initialization of genetic algorithm** In the genetic algorithm, the population is the number of chromosomes in the search space and we selected a population size of 40. Since GA starts with no chromosome at first pass, the initial population is generated from random way.

**Fitness evaluation** The purpose of the designed algorithm is to select the most relevant features that produce the highest classification accuracy for the

diagnosis of PD. In the genetic algorithm, the eligibility of the chromosomes to be selected depends on their order given by the fitness function. Our algorithm uses SVM as a formatting function to classify chromosomes (subset of features) at each turn.

**Genetic operators** At each generation, the algorithm uses three genetic operators, i.e. selection, mutation and crossover. Selection process depends on the fitness of the chromosomes. However, the most eligible chromosomes must be retained in the population at each turn. The proposed algorithm provides this with the use of an elitist roulette wheel selection scheme [42]. Crossover makes use of a substring exchange between two chromosomes to generate two new chromosomes. The simplest form of crossover is known as single point crossover. Other types are two point crossover, Uniform crossover [43]. In mutation, as a last genetic operator, selected genes are inverted preventing search process not to get stuck in local maxima. Usually the commonly suggested crossover rate would tend to be fairly average (around 0.5), while the mutation rate is low (typically 0.01 to 0.15) for efficient research [44]. In the proposed model, we use single point crossover with probability of 0.5 and the mutation probability is selected to be 0.04.

**Algorithm 1** Genetic algorithm with SVM classifier for feature selection.

```
Data normalization
Split data into k=10 folds
for k = 1tok=10 do
    Specify
    A = test data (1 fold reserved for SVMclassifier)
    B = train data (8 folds train for SVMcore)
    C = validation data (1 fold validation for SVMcore)
    repeat
        for train and validation data
        Step 1: Encode features as binary chromosomes
        Step 2: Generate randomly a population of 40 chromosomes
        Step 3: Evaluate accuracy of SVMcore algorithm for step 2
        Step 4: Apply binary cross-over with probability of 0.5
        Step 5: Apply binary mutation with probability of 0.04
        Step 6: Calculate new SVMcore accuracy of chromosomes and compare it with
        step 3
        Step 7: Select best chromosomes with highest fitness
        Step 8: Replace chromosomes with lowest fitness
    until (up to 30 generations)
    Train SVMclassifier with data (B + C)
    Test SVMclassifier with test data (A)
    Calculate predictive accuracy for fold k
end for
Calculate average predictive accuracy for 10 folds
```

The proposed algorithm generates new populations by GA operators and fitness of chromosomes in each population is calculated to rank them. In the designed algorithm, a termination criterion for this ongoing process is the number of generations. The algorithm is stopped when the number of generations is equal to 30. **Algorithm 1** shows the operational steps of the proposed method Furthermore, it must be noted in **Algorithm 1** that there are two SVM algorithms; (i) SVMcore as fitness evaluator of GA and (ii) SVMclassifier as the evaluator of feature model generated by GASVMcore. The designed algorithm has two main loops attached to each other: (i) with the k-fold cross validation approach, the data is divided into 10 folds. In the GA-SVM loop, once one fold is used as a test set, the remaining 8 are held for training and the last for validation.

The GA-SVM combination is run over 30 generations and the most important features are obtained in this loop. (ii) In the second loop, the validation set and 8 train sets with the new optimized features of the loop (i) are combined as a training set for the SVMcore algorithm. After trained the SVM algorithm with this combined train set, the algorithm is used to classify the instances of the reserved test set of the loop (i). This process is repeated 10 times to obtain average classification accuracy.

As mentioned earlier, different learning methods and eight feature selection techniques along with SVM classifier were used to evaluate the effectiveness of our proposed algorithm, that are t-test [36], Info-gain, CFS, RFS, Relieff, MRMR, GA and wrapper algorithms based on SVM-RFE-CBR [25].

### 3.2 Support Vector Machine

The SVM has been one of the more widely used data learning tools in recent years. It is usually used to address a binary pattern classification problem. The binary SVM constructs a set of hyper-plan in an infinite dimensional space, which can then be divided into two kinds of representations, such as the linear and nonlinear SVM [37]. In particular, the linear SVM is given by:

$$f(x) = w^T x - b \tag{1}$$

Where x is the data, y is the label of the class, w is the weight vector orthogonal to the decision hyper-plane, b the offset of the hyper-plan and T is the transposition operator. The solution to linear SVM is found by maximizing the margin used to separate classes. This is equivalent to solving the following minimization problem:

$$min\left\{\frac{1}{2}w^T w + c\sum_{i=1}^{n}\epsilon_i\right\} \tag{2}$$

Subject to,

$$y_i(w^T x_i + b) \geq (1 - \epsilon_i) \tag{3}$$

Where

$$\epsilon(\epsilon_i \geq, i = 1, 2, ..., n)$$

is a slack variable used to indicate the degree of classification error allowed, $C > 0$ is a penalty parameter corresponding to the upper limit of the error, and $n$ is the number of instances. The non-linear SVM classifier uses a K-function of the kernel to separate data nonlinearly. It expresses as follows:

$$f(x_i) = sign\left(\sum_{i=1}^{n} y_i \alpha_i K\left\langle x, x_i \right\rangle + b\right) \tag{4}$$

Where $\alpha$ is the Lagrange multiplier,$K$ is a function of the kernel and $b$ is a constant coefficient. In our work, we adopt the radial base function (RBF) as a nonlinear kernel. It is given by:

$$K(x, x_i) = exp\left(\frac{||x_i - x||}{2\sigma^2}\right) \tag{5}$$

Where $\sigma > 0$ is a scale parameter [37]. Since the choice of kernel optimised parameters remains delicate, In our method, the value of the slack variable is set to 0.001. for each selection along with each run, we optimize the penalty parameter $c$ and the scale parameter $\sigma$ of the cross-validated SVM classifier using Bayesian optimization to decide on the hyperparameters to optimize [38].

### 3.3 Performance measures and cross-validation protocol

Finally, classification precision, sensitivity, accuracy, f-score and area under curve (Auc) are all employed to assess the effectiveness of each machine learning classifier in distinguishing between healthy normal (HN) and PD patients [21]. They are expressed as follows:

$$Precision = \frac{TP}{FP + FN} \tag{6}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{7}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{8}$$

$$F - score = \frac{2TP}{2TP + FP + FN} \tag{9}$$

$$Auc = \int_{-\infty}^{+\infty} TPrate(t) FPrate(t)\, \mathrm{d}t \tag{10}$$

Where $TP, TN, FPandFN$ designate true positives (i.e., PD patients), true negatives (i.e., a healthy normal), false positives and false negatives, respectively. Finally,

$$TPrate = TP/nP$$

and

$$FPrate = FP/nP$$

, where $nP$ is the number of positive samples (PD patients) and $t$ is a variable parameter in $[0, 1]$ [32].
For the robustness of the performance evaluation, a k-fold cross validation protocol is used in this work. In particular, we adopted the ten-fold cross-validation in which nine subsets are used as a training set, while the remaining one is used to test the trained model. Then, the cross-validation process is repeated ten times, with each of the ten subsets being used exactly once for validation. At each repetition, each classification performance measure is calculated. Then, its associated mean and standard deviation over ten repetitions are calculated [31].

## 4 Experiment details

### 4.1 Dataset

The Parkinson dataset used in this study is composed of a range of biomedical voice measurements from 31 people, 23 with Parkinson disease (PD), and it includes a total of 195 voice recordings from individuals. In addition, these biomedical voice measurements have different feature information given in **Table 1** The full description of these features can be found in [18].

### 4.2 Evaluation of the performance

In this paper, the modelling and evaluation operations are implemented in a Python development environment, where the packages of the Scikit-learn [41], the Skfeature [40] and the Hyperopt [39] are employed for computational verification of the baseline learning methods, the feature selection algorithms and the Bayesian hyper-parameter optimization, respectively. The results of the proposed method are compared on various performance measures indicated previously. Following the ten-fold cross-validation, the average and standard deviation of each performance measure were calculated. **Table 2** presents the comparison of the proposed method with the results of the other learning methods such as Neural Network, SVM and KNN on the average accuracy, precision, recall and the area under the curve (Auc). According to this table, the proposed method achieves the highest results for all measurements compared to other learning methods. The second best classification results relate to SVM. In addition, ANN outperformed the other methods on both Precision and Auc. The performance measures of the KNN and DT approaches are close to each other.

**Table 4** shows the influence of different feature selection methods on the classification performance of SVM in which the priority of the proposed GA with SVM to the other feature selection methods is clear. As shown in **Table 3** , the feature space is varied on the basis of different feature selection algorithms

**Table 1.** The features of biomedical voice measurements.

| Number | Attributes | Explanation |
|---|---|---|
| 0 | MDVP:Fo (Hz) | Average vocal fundamental frequency |
| 1 | MDVP:Fhi (Hz) | Maximum vocal fundamental frequency |
| 2 | MDVP:Flo (Hz) | Minimum vocal fundamental frequency |
| 3 | MDVP:Jitter (%) | Several measures of variation in undamental frequency |
| 4 | MDVP:Jitter (abs) | |
| 5 | MDVP:RAP | |
| 6 | MDVP:PPQ | |
| 7 | Jitter:DDP | |
| 8 | MDVP:Shimmer | Several measures of variation in amplitude |
| 9 | MDVP:Shimmer (dB) | |
| 10 | Shimmer:APQ3 | |
| 11 | Shimmer:APQ5 | |
| 12 | MDVP:APQ | |
| 13 | Shimmer:DDA | |
| 14 | RPDE | Two nonlinear dynamical complexity measures |
| 15 | D2 | |
| 16 | NHR | The measure of ratio of noise to tonal components in the voice status |
| 17 | HNR | |
| 18 | DFA | Signal fractal scaling exponent |
| 19 | spread1 | Three nonlinear measures of fundamental frequency variation |
| 20 | spread2 | |
| 21 | PPE | |

**Table 2.** Comparisons of all methods on different performance measures.

| S. no | Classification algorithm | Accuracy | Precision | Recall | F-score | Auc |
|---|---|---|---|---|---|---|
| 1 | SVM | 0.88±0.07 | 0.88±0.07 | 0.97±0.03 | 0.92±0.04 | 0.78±0.14 |
| 2 | ANN | 0.86±0.08 | 0.89±0.08 | 0.93±0.05 | 0.91±0.05 | 0.79±0.15 |
| 4 | Decision trees | 0.81±0.07 | 0.88±0.06 | 0.87±0.07 | 0.87±0.05 | 0.75±0.10 |
| 3 | KNN | 0.82±0.08 | 0.87±0.07 | 0.89±0.06 | 0.88±0.05 | 0.74±0.13 |
| 5 | Proposed method | 0.95±0.05 | 0.96±0.05 | 0.98±0.04 | 0.96±0.03 | 0.92±0.09 |

**Table 3.** Comparisons of feature space selection algorithms.

| Feature selection algorithms | Selected features |
|---|---|
| Info Gain | 03 (0, 18, 21) |
| T-test | 04 (21, 18, 12, 8) |
| CFS | 12 (13, 19, 0, 1, 2, 4, 15, 16, 17, 18, 20, 21) |
| RFS (p-value) | 12 (3, 7, 5, 11, 4, 8, 18, 0, 9, 16, 14, 15) |
| Relief | 06 (0, 18, 17, 21, 19, 16) |
| MRMR | 16 (0, 4, 9, 5, 6, 3, 10, 7, 13, 12, 14, 8, 11, 19, 1, 2) |
| RFE-CBR | 12 (0, 2, 3, 4, 5, 7, 8, 11, 14, 15, 16, 18) |
| Genetic algorithm | 10 (0, 7, 11, 12, 13, 14, 15, 16, 18, 21) |
| Genetic algorithm with SVM | 09 (0, 7, 10, 12, 13, 14, 15, 16, 21) |

(**Table 4** ). Also, Figures .35 present box plots of accuracy, precision, and recall distributions for all feature selection algorithms under study. According to the accuracy distributions shown in **Fig. 3** , GA-SVM, GA and RFE-CBR selection techniques are more suitable for SVM than other techniques. In addition, feature selection techniques based on CFS and Information Gain offer lower accuracy than other techniques. By inspecting the precision distributions in **Fig. 4** , GA-SVM and RFE-CBR based feature selection techniques work better than other techniques. The second best precision concerns the selection techniques RFS, Relieff and GA. In addition, T-test and CFS offer lower precision than other selection methods. Finally, by examining the recall distributions of **Fig. 5** , it can be seen that the GA-SVM and GA feature selection techniques have high recall results. In contrast, the MRMR method provides lower recall values.

**Table 4.** Comparisons of the classification results of svm in the reduced feature space and full feature set.

| Method | #Selected feature | The average and standard deviation of each performance measure. | | | | |
|---|---|---|---|---|---|---|
| | | Accuracy | Precision | Recall | F-score | Auc |
| Full feature set | 22 | 0.886± 0.07 | 0.88 ± 0.07 | 0.97±0.03 | 0.92 ± 0.04 | 0.78 ± 0.14 |
| Info Gain | 03 | 0.880 ± 0.11 | 0.906 ± 0.09 | 0.945 ± 0.11 | 0.924 ± 0.07 | 0.810 ± 0.19 |
| T-test | 04 | 0.895 ± 0.08 | 0.890 ± 0.09 | 0.946 ± 0.07 | 0.912 ± 0.06 | 0.770 ± 0.18 |
| CFS | 12 | 0.871 ± 0.07 | 0.902 ± 0.07 | 0.938 ± 0.05 | 0.917 ± 0.04 | 0.801 ± 0.13 |
| Relieff | 06 | 0.895 ± 0.08 | 0.935 ± 0.08 | 0.937 ± 0.11 | 0.928 ± 0.06 | 0.853 ± 0.11 |
| MRMR | 16 | 0.871 ± 0.11 | 0.912 ± 0.08 | 0.924 ± 0.09 | 0.915 ± 0.07 | 0.812 ± 0.17 |
| RFE-CBR | 12 | 0.906 ± 0.06 | 0.943 ± 0.07 | 0.939 ± 0.03 | 0.939 ± 0.04 | 0.872 ± 0.12 |
| GA | 10 | 0.926 ± 0.07 | 0.935 ± 0.06 | 0.972 ± 0.04 | 0.953 ± 0.05 | 0.878 ± 0.11 |
| GA-SVM | 09 | 0.953 ± 0.05 | 0.962 ± 0.05 | 0.980 ± 0.04 | 0.969 ± 0.03 | 0.927 ± 0.09 |

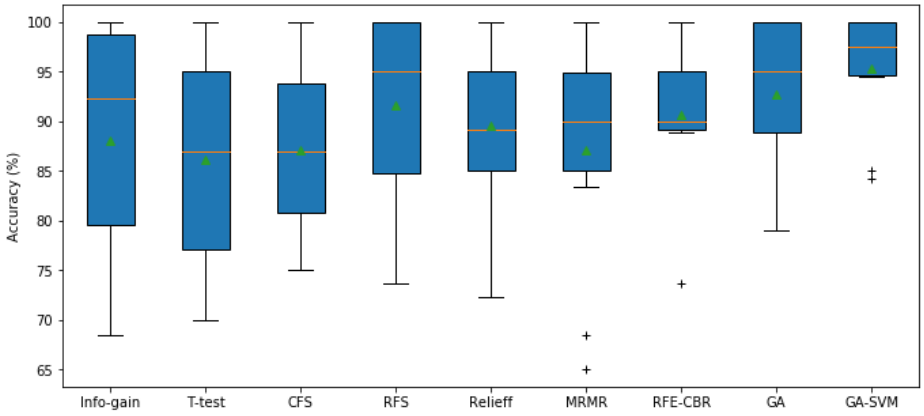


**Fig. 3.** Boxplot of the SVM accuracy on reduced feature space of each selection techniques. The symbol + indicates an outlier

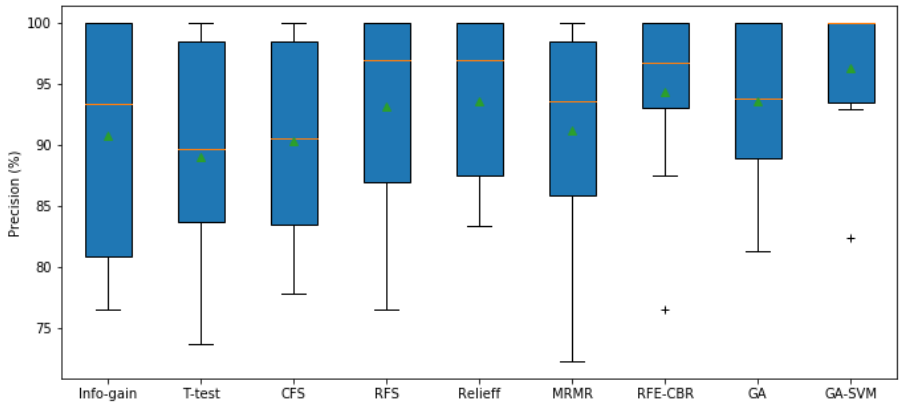


**Fig. 4.** Boxplot of the svm precision on reduced feature space of each selection techniques. The symbol + indicates an outlier

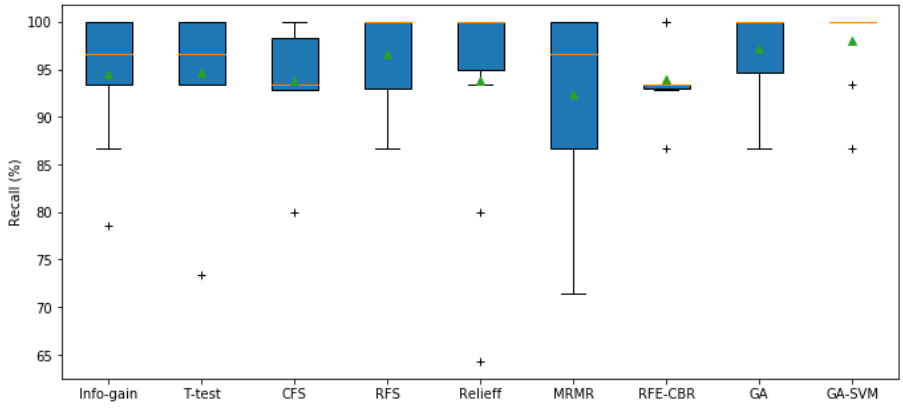


**Fig. 5.** boxplot of the SVM recall on reduced feature space of each selection techniques. The symbol + indicates an outlier.

## 5 Conclusion

A genetic approach wrapped Support vector machine is proposed in this paper for improving the performance of GA. GA-SVM based feature selection algorithms identifies 9 features (MDVP:Fo (Hz), Jitter:DDP, Shimmer:APQ3, MDVP:APQ, Shimmer:DDA, RPDE, D2, NHR, and PPE) to detect Parkinson's disease. The resulting functions are provided to SVM to search for accuracy. This methodology outperformed other learning method as a high performance diagnosis results are showed consist of 0.95, 0.96, 0.98, 0.92 for accuracy, precision, recall and area under curve. Also the application of feature selection algorithms after SVM shows the improvement in performance. further our method perform better than various feature selection methods including Information gain, T-test, CFS, RFS (R-value), relieff, MRMR, RFE-CBR and GA.

## References

1. Harel, B.-T., Cannizzaro, .S, Cohen, H., Reilly ,N., Snyder ,P.-J.: Acoustic characteristics of Parkinsonian speech: A potential biomarker of early disease progression and treatment. Journal of Neurolinguistics **17**(6), 439–453 (2004)
2. Singh, N., Pillay, V., Choonara, Y.-E.: Advances in the treatment of Parkinsons disease. Progress in Neurobiology **81**(1), 29–44 (2007)
3. Jankovic, J.:Parkinsons disease: clinical features and diagnosis. Journal of Neurology, Neurosurgery & Psychiatry **79**, 368–376 (2008)
4. Massano, J., Bhatia, K.-P.: Clinical approach to Parkinsons disease: features, diagnosis, and principles of management. Cold Spring Harbor Perspectives in Medicine **2**(6), (2012)
5. Lan, K.-C., Shih, V.-Y.: Early Diagnosis of Parkinsons Disease Using a Smartphone. Procedia Comput. Sci. **34**, 305–312 (2014)

6. Ma, C., Ouyang, J., Chen, H.-L., Zhao, X.-H.: An efficient diagnosis system for Parkinsons disease using kernel-based extreme learning machine with subtractive clustering features weighting approach. Comput Math Methods Med **2014**, 1–14 (2014)
7. Hossen, A., Muthuraman, M., Raethjen, J., Deuschl, G., Heute, U.: Discrimination of Parkinsonian tremor from essential tremor by implementation of a wavelet-based soft-decision technique on EMG and accelerometer signals. Biomed Signal Process Control **5**, 18 (2010)
8. Daliri, M.-R.: Chi square distance kernel of the gaits for the diagnosis of Parkinsons disease. Biomed Signal Process Control **8**(1), 66–70 (2013)
9. Duffy, R.-J. Motor Speech Disorders: Substrates, Differential Diagnosis and Management. 2nd Edition, Elsevier Mosby, St. Louis. (2005)
10. Ho, A.-K., Lansek, R., Marigliani, C., Bradshaw, J.-L., Gates, S.: Speech Impairment in a Large Sample of Patients with Parkinsons Disease. Behaviour Neurology **11**, 131–137 (1998)
11. Sapir, S., Spielman, J.-L., Ramig, L.-O., Story, B.-H., Fox, C.: Effects of intensive voice treatment (the Lee Silverman voice treatment [LSVT]) on vowel articulation in dysarthric individuals with idiopathic Parkinson disease: Acoustic and perceptual findings. J. Speech. Lang. Hear. Res. **50**, 899–912 (2007)
12. Rahn, D.-A., Chou, M., Jiang, J.-J., Zhang, Y.: Phonatory Impairment in Parkinsons Disease: Evidence from Nonlinear Dynamic Analysis and Perturbation Analysis. Journal of Voice **21** 64–71 (2007)
13. Peng, B., Wang, S., Zhou, Z., Liu, Y., Tong, B., Zhang, T., Dai, Y.: A multilevel-ROI-features-based machine learning method for detection of morphometric biomarkers in Parkinsons disease. Neurosci. Lett. **651**, 88–94 (2017)
14. Salvatore, C., Cerasa, A., Castiglioni, I., Gallivanone, F., Augimeri, A., Lopezd, M., Arabia, G., Morelli, M., Gilardi, M.-C., Quattrone A.: Machine learning on brain MRIdata for differential diagnosis of Parkinsons disease and ProgressiveSupranuclear. Palsy, J. Neurosci. Methods **222**, 230–237 (2014)
15. Yang, S., Zheng, F., Luo, X., Cai, S., Wu, Y., Liu, K., Wu, M., Chen, Krishnan, J.-S: Effective dysphonia detection using feature dimension reduction and kernel density estimation for patients with Parkinsons disease. PLoS One **9**(2):e88825 (2014)
16. Kotsavasiloglou, C., Kostikis, N., Hristu-Varsakelis, D., Arnaoutoglou M.: Machine learning-based classification of simple drawing movements in Parkinsons disease. Biomed. Signal Process. Control **31** 174–180 (2017)
17. Kaya, E., Findik, O., Babaoglu, I., Arslan, A. Effect of discretization method on the diagnosis of Parkinsons disease. International Journal of Innovative Computing, Information and Control **7** 4669–4678 (2011)
18. Tsanas, A., Little, M.-A., McSharry, P.-E., Spielman, J., Ramig, L.-O.: Novel speechsignal processing algorithms for high-accuracy classification of Parkinsonsdisease. IEEE Trans. Biomed. Eng. **59**, 1264–1271 (2012)
19. Sakar, B.-E., Isenkul, M.-E., Sakar, C.-O., Sertbas, A., Gurgen, F., Delil, S., Apaydin, H., Kursun, O.: Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. IEEE J Biomed Health Inform **17**(4), 828-834 (2013)
20. Can, M.: Neural networks to diagnose the Parkinsons disease. South East Europe Journal of Soft Computing **2**(1), (2013)
21. Shahsavari, M.-K., Rashidi, H., Bakhsh H.-R.: Efficient classification of Parkinsons disease using extreme learning machine and hybrid particle swarm optimization. In: 2016 4th International Conference on Control, Instrumentation, and Automation (ICCIA), pp. 148–154. IEEE, Qazvin (2016)

22. Visalakshi, S., Radha, V.: A literature review of feature selection techniques and applications: Review of feature selection in data mining. In: IEEE International Conference on Computational Intelligence and Computing Research (ICCIC) pp. 1–6. (2014)
23. Yana, K., Zhang, D.: Feature selection and analysis on correlated gas sensor datawith recursive feature elimination, Sens. Actuators B: Chem. **212**, 353-363 (2015)
24. Shahbaba, B., Neal, R.: Nonlinear models using Dirichlet process mixtures. Journal of Machine Learning Research, **10**, 1829-1850 (2009)
25. Little, M.-A., McSharry, P.-E., Hunter, E.-J., Ramig L.-O.: Suitability of dysphonia measurements for telemonitoring of Parkinsons disease. IEEE Transactions on Biomedical Engineering **56**, 1015-1022 (2009)
26. Das, R. : A comparison of multiple classification methods for diagnosis of Parkinson disease. Expert Systems with Applications **37** (2), 1568-1572 (2010)
27. Guo, P.-F., Bhattacharya, P., Kharma, N.: Advances in detecting Parkinsons disease. Medical Biometrics **6165**, 306–314 (2010)
28. Ozcift, A., Gulten, A.: Classifier ensemble construction with rotation forest to improve medical diagnosis performance of machine learning algorithms. Computer Methods and Programs in Biomedicine **104**(3), 443-451 (2011)
29. Chen, H.-L., Huang, C.-C., Yu, X. G.: An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Systems with Applications **40**(1), 263–271 (2013)
30. Chen, H.-L., Wang, G., Ma, C., Cai, Z.-N., Liu, W.-B., Wang, S.-J.: An efficient hybrid kernel extreme learning machine approach for early diagnosis of Parkinson's disease. Neurocomputing **184**(4745), 131–144 (2016)
31. Peker, M., Sen, B., Delen, D.: Computer-Aided Diagnosis of Parkinson's Disease Using Complex-Valued Neural Networks and mRMR Feature Selection Algorithm. J Healthc Eng. **6**(3), 281–302 (2015)
32. Lahmiri, S., Shmuel, A.: Detection of Parkinsons disease based on voice patterns ranking and optimized support vector machine. Biomed. Signal Process Control **49 2019**, 427–433 (2019)
33. Goldberg, D.-E.: Genetic algorithms in search, optimization and machine learning. Addison-Wesley, New York (1998)
34. Gokulnath, C.-B., Shantharajah, S.-P.: An optimized feature selection based on genetic approach and support vector machine for heart disease. Cluster Comput 1–11 (2018)
35. Yan, H., Zheng, J., Jiang, Y., Peng, C., Xiao, S.: Selecting critical clinical features for heart diseases diagnosis with a real-coded genetic algorithm. Appl. Soft Comput **8**, 1105–1111 (2008)
36. Fayyad, U., Irani, K.: Multi-interval discretization of continuous valued attributes for classification learning. in: Proceeding of The International JointConference on Artificial Intelligence, pp. 1022-1029 (1993)
37. Chao, C.-F. Horng M.-H.: The construction of support vector machine classifier using the firefly algorithm. Comput Intell Neurosci **2015**, 8 (2015)
38. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning, DataMining, Inference, and Prediction. Second edition. Springer Verlag, New York (2009)
39. Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., Cox, D.-D.: Hyperopt: a Python library for model selection and hyperparameter optimization. Computational Science & Discovery **8**(1), 1749–4699 (2015)

40. Li, J.-D, Cheng, K.-W., Wang, S.-H., Morstatter, F., Trevino, R.,-P., Tang, J.-L, Liu, H.: Feature selection: A data perspective. ACM Computing Surveys (CSUR) **50**(6), 94 (2018)
41. Pedregosa et al: Scikit-learn: Machine Learning in Python. JMLR **12**, 2825–2830 (2011)
42. Abd-Rahman, R., Ramli, R., Jamari, Z., Ku-Mahamud, K.-R.: Evolutionary algorithm with roulette-tournament selection for solving aquaculture diet formulation. Math Probl Eng **2016**, 1–10 (2016)
43. Jadhav, S., He, H.-M.: Information gain directed genetic algorithm wrapper feature selection for credit rating. Appl Soft Comput **69**, 541553 (2018)
44. Avci, D., Dogantekin, A.: An Expert diagnosis system for Parkinson disease based on genetic algorithm-wavelet kernel-extreme learning machine. Parkinsons disease 19 (2016)